

Folha de Pagamento - SPOBRAS 2019

```
library(tidyverse)
library(na.tools)
library(readxl)
library(stringr)
library(lubridate)
library(pdftools)
library(data.table)
```

```
spobras<-fread("http://dados.prefeitura.sp.gov.br/dataset/3edd72cc-2f79-4c95-88e7-55744881b0a6/resource/
```

Folha de pagamento SPOBRAS - 2019

```
spobras<-as_tibble(spobras)
spobras
```

```
## # A tibble: 2,362 x 15
##   EMPRESA MES      ANO NOME  CARGO LOTAÇÃO ADMISSÃO NASCIMENTO VENCIMENTOS
##   <chr>    <chr> <int> <chr> <chr> <chr>    <chr>          <int> <chr>
## 1 SPObras Jane~ 2019 WALT~ ANAL~ NUCLEO~ 19/08/1~      1960 13482,77
## 2 SPObras Jane~ 2019 LUIZ~ AGEN~ NUCLEO~ 12/10/1~      1962 0
## 3 SPObras Jane~ 2019 VALD~ AGEN~ NUCLEO~ 10/04/1~      1965 6622,49
## 4 SPObras Jane~ 2019 MARC~ ANAL~ NUCLEO~ 14/07/1~      1959 23401,38
## 5 SPObras Jane~ 2019 MARC~ AGEN~ NUCLEO~ 03/08/1~      1956 6339,57
## 6 SPObras Jane~ 2019 JOAO~ AGEN~ NUCLEO~ 17/09/1~      1959 5255,93
## 7 SPObras Jane~ 2019 ELIA~ SECR~ GESTAO~ 01/12/1~      1955 9797,86
## 8 SPObras Jane~ 2019 SILV~ ANAL~ NUCLEO~ 28/12/1~      1962 14174,82
## 9 SPObras Jane~ 2019 MARC~ ANAL~ NUCLEO~ 01/11/1~      1958 14174,82
## 10 SPObras Jane~ 2019 ALFR~ AGEN~ NUCLEO~ 18/11/1~      1964 13508,67
## # ... with 2,352 more rows, and 6 more variables: ENCARGOS <chr>,
## #   BENEFÍCIOS <chr>, 'OUTRAS REMUNERAÇÕES' <int>, VÍNCULO <chr>, 'DETALHE
## #   VÍNCULO' <chr>, LIMINAR <chr>
```

CONTAGEM DE VALORES FALTANTES EM NOSSO DATASET.

```
sum(is.na(spobras))
```

```
## [1] 0
```

EXCLUINDO A VARIÁVEL LIMINAR

```
spobras<-spobras[,-15]
```

QUANTIDADE DE SERVIDORES EM NOSSO DATASET SPOBRAS

```
spobras %>% distinct(NOME) # RETORNA 241 NOMES
```

```
## # A tibble: 241 x 1
##   NOME
##   <chr>
## 1 WALTER MARTINS DE SOUZA
## 2 LUIZ CARLOS DE SOUZA BARBOZA
## 3 VALDELI DA SILVA
## 4 MARCIA APARECIDA TEIXEIRA DE CAMPOS CARVALHO
## 5 MARCOS SILVA ROCHA
## 6 JOAO PAULO RESPLANDES DE OLIVEIRA
## 7 ELIANA MARIANO FRANCO
## 8 SILVANA ISGROI CARVALHO
## 9 MARCIA HELENA NUCCI EIRAS GARCIA
## 10 ALFREDO LENZ
## # ... with 231 more rows
```

QUANTIDADE DE CARGOS DIFERENTES NO DATASET

```
spobras %>% distinct(CARGO) # RETORNA 44 CARGOS DIFERENTES
```

```
## # A tibble: 44 x 1
##   CARGO
##   <chr>
## 1 ANALISTA DE PROJETOS, OBRAS E SERVIÇOS-ENGENHEIRO
## 2 AGENTE DE PROJETOS, OBRAS E SERVIÇOS
## 3 AGENTE DE ADMINISTRAÇÃO
## 4 ANALISTA DE ADMINISTRAÇÃO
## 5 SECRETARIA
## 6 ASSISTENTE DE PROJETOS OBRAS E SERVIÇOS
## 7 ANALISTA DE AGRIMENSURA E CONSTRUÇÃO CIVIL
## 8 AUXILIAR DE ADMINISTRAÇÃO
## 9 ANALISTA DE PROJETOS OBRAS E SERVIÇOS - ARQUITETO
## 10 ADVOGADO
## # ... with 34 more rows
```

IDENTIFICANDO A QUANTIDADE DE VINCULOS NO DATASET

```
spobras %>% group_by('DETALHE VÍNCULO') %>% count('DETALHE VÍNCULO')
```

```
## # A tibble: 4 x 2
## # Groups:   DETALHE VÍNCULO [4]
##   'DETALHE VÍNCULO'      n
##   <chr>                <int>
## 1 Concursado NÃO ocupante de cargo em comissão    1371
## 2 Concursado ocupante de cargo em comissão         43
## 3 Não concursado NÃO ocupante de cargo em comissão  444
## 4 Não concursado ocupante de cargo em comissão    504
```

TRANSFORMANDO VARIÁVEIS VENCIMENTOS, ENCARGOS E BENEFÍCIOS EM VARIÁVEIS NUMÉRICAS

```
### VARIÁVEIS: VENCIMENTOS, ENCARGOS E BENEFÍCIOS ESTÃO COMO strings.. É PRECISO CONVERTE-LAS A NUMEROS
## (substituir a vírgula por ponto)
spobras$VENCIMENTOS<-str_replace_all(spobras$VENCIMENTOS,pattern = "[,]",replacement = ".")
spobras$ENCARGOS<-str_replace_all(spobras$ENCARGOS,pattern = "[,]",replacement = ".")
spobras$BENEFÍCIOS<- str_replace_all(spobras$BENEFÍCIOS,pattern = "[,]",replacement = ".")

### agora transforma-las em numeros com casas decimais
spobras$VENCIMENTOS<-as.double(spobras$VENCIMENTOS)
spobras$ENCARGOS<-as.double(spobras$ENCARGOS)
spobras$BENEFÍCIOS<-as.double(spobras$BENEFÍCIOS)
```

SEPARANDO DATASET POR TIPO DE VÍNCULO EMPREGATÍCIO

```
cncm<-spobras %>% select(NOME,CARGO,VENCIMENTOS,'DETALHE VÍNCULO') %>%
  filter('DETALHE VÍNCULO' == "Concursado NÃO ocupante de cargo em comissão") ## 1371 OBSERVAÇÕES

ccc<-spobras %>% select(NOME,CARGO,VENCIMENTOS,'DETALHE VÍNCULO') %>%
  filter('DETALHE VÍNCULO' == "Concursado ocupante de cargo em comissão") ## 43 OBSERVAÇÕES

ncnc<-spobras %>% select(NOME,CARGO,VENCIMENTOS,'DETALHE VÍNCULO') %>%
  filter('DETALHE VÍNCULO' == "Não concursado NÃO ocupante de cargo em comissão") # 504 OBSERVAÇÕES

nccc<-spobras %>% select(NOME,CARGO,VENCIMENTOS,'DETALHE VÍNCULO') %>%
  filter('DETALHE VÍNCULO' == "Não concursado ocupante de cargo em comissão") ## 444 OBSERVAÇÕES
```

DISTRIBUIÇÃO DE SERVIDORES POR VÍNCULOS E SALÁRIOS

```
# PROPORÇÃO DE SERVIDORES POR VÍNCULO
proporcao_servidores<-data.frame(Vinculo=c("Concursado Não Ocupante CM","Concursado Ocupante CM","Não O
  Qtd_servidores=rbind(nrow(cncm),nrow(ccc),nrow(ncnc),nrow(nccc),nrow(spobras)),

### PROPORÇÃO SALÁRIO BRUTO POR VÍNCULO
proporcao_salarial_X_vinculo<-data.frame(Vinculo=c("Concursado Não Ocupante CM","Concursado Ocupante CM
  Salario_Bruto_total=rbind(sum(cncm$VENCIMENTOS),sum(ccc$VENCIMENTOS),sum(ncnc$VENCIMENTOS),sum(nccc$VEN
  Proprcao_salario_Bruto=rbind(sum(cncm$VENCIMENTOS)/sum(spobras$VENCIMENTOS)*100, sum(ccc$
```

TRANSFORMANDO EM TIBBLES

```
proporcao_servidores<-as_tibble(proporcao_servidores)
proporcao_salarial_X_vinculo<-as_tibble(proporcao_salarial_X_vinculo)
```

TABELA E GRÁFICO: PROPORÇÃO SERVIDORES POR VÍNCULO

```
proporcao_servidores
```

```
## # A tibble: 5 x 3
##   Vinculo                Qtd_servidores Proporcao
##   <chr>                  <int>      <dbl>
## 1 Concursado Não Ocupante CM      1371      58.0
## 2 Concursado Ocupante CM           43       1.82
## 3 Não Concursado Não Ocupante CM    444      18.8
## 4 Não Concursado Ocupante CM       504      21.3
## 5 Total                        2362     100
```

```
proporcao_servidores %>%
  ggplot(mapping = aes(x=Proporcao,y=Vinculo,fill=Vinculo))+
    geom_col(show.legend = FALSE)+
    ggtitle(label = "Proporção Servidores por Vínculo Empregatício")+
    xlab(label = "Porcentagem (%)")+
    ylab(label = "Vínculo Empregatício")
```

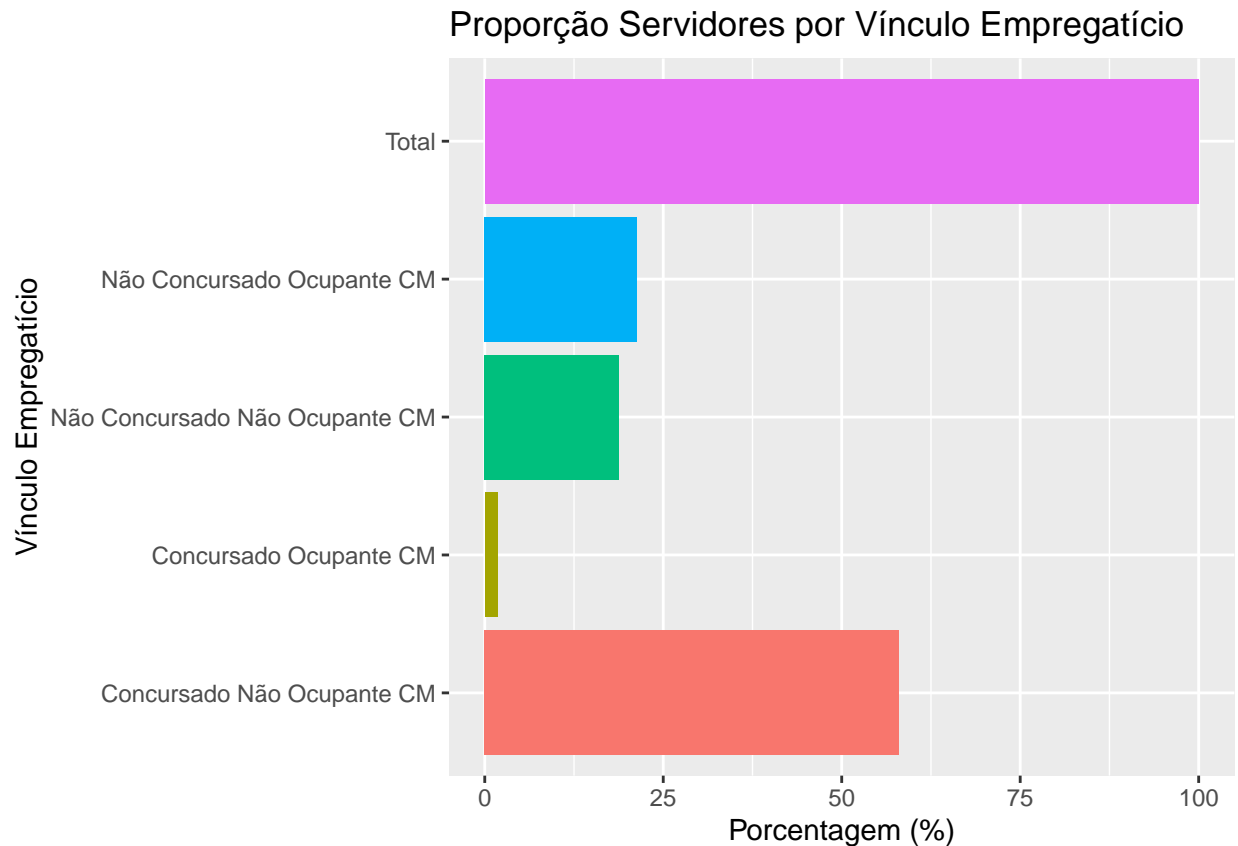


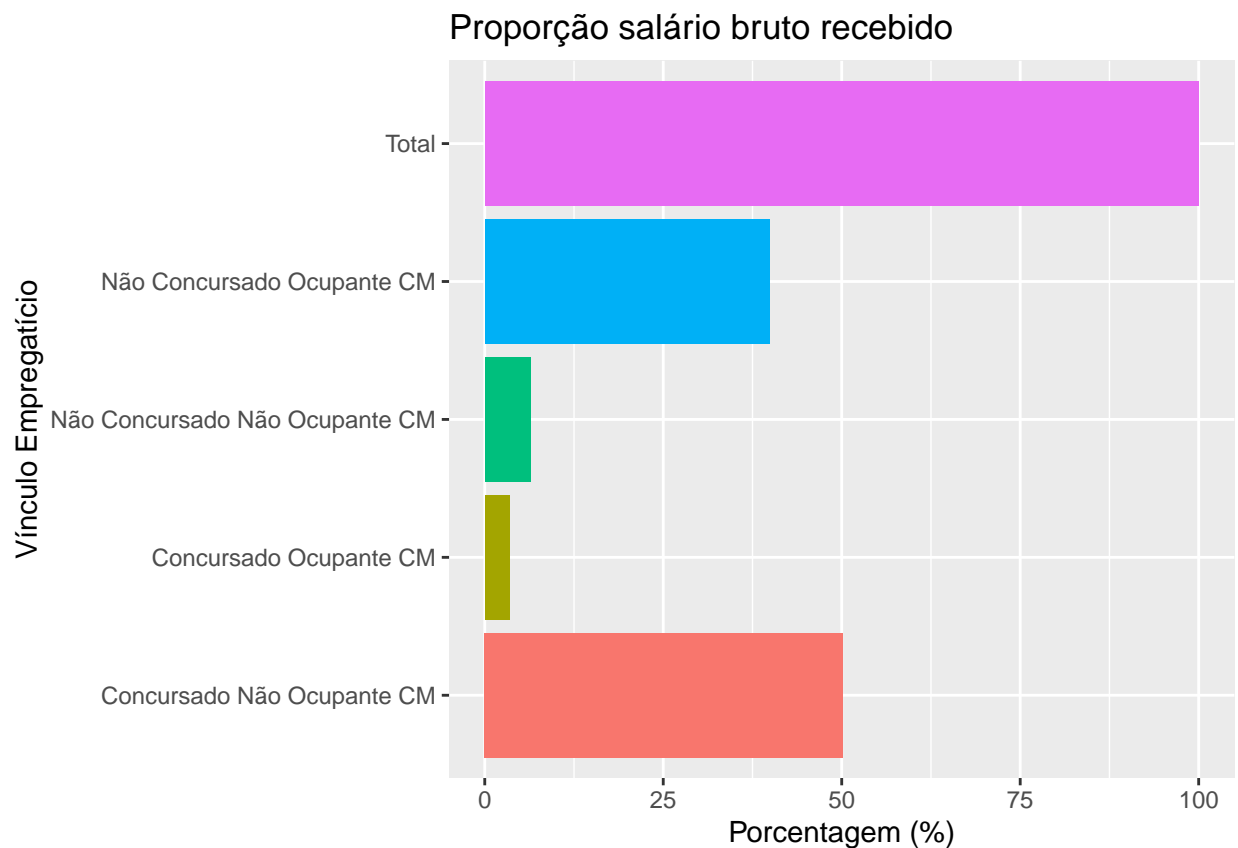
TABELA E GRAFICO: PROPORÇÃO SERVIDORES SALARIO BRUTO POR VINCULO

```
proporcao_salarial_X_vinculo
```

```
## # A tibble: 5 x 3
##   Vinculo                Salario_Bruto_total Proprcao_salario_Bruto
```

##	<chr>	<dbl>	<dbl>
## 1	Concursado Não Ocupante CM	11042321.	50.2
## 2	Concursado Ocupante CM	768121.	3.49
## 3	Não Concursado Não Ocupante CM	1416981.	6.44
## 4	Não Concursado Ocupante CM	8779982.	39.9
## 5	Total	22007405.	100

```
proporcao_salarial_X_vinculo %>%
  ggplot(mapping = aes(x=Proprcao_salario_Bruto,y=Vinculo,fill=Vinculo))+
  geom_col(show.legend = FALSE)+
  ggtitle(label = "Proporção salário bruto recebido")+
  xlab(label = "Porcentagem (%)")+
  ylab(label = "Vínculo Empregatício" )
```



TRANSFORMANDO DATA DE ADMISSÃO EM DATA RECONHECIDA PELO R

```
spobras$ADMISSÃO<-dmy(spobras$ADMISSÃO)
spobras %>% select(ADMISSÃO)
```

```
## # A tibble: 2,362 x 1
##   ADMISSÃO
##   <date>
```

```
## 1 1976-08-19
## 2 1979-10-12
## 3 1980-04-10
## 4 1980-07-14
## 5 1981-08-03
## 6 1981-09-17
## 7 1981-12-01
## 8 1981-12-28
## 9 1983-11-01
## 10 1987-11-18
## # ... with 2,352 more rows
```

CRIANDO UMA VARIÁVEL COM OS ANOS DE TRABALHO DE CADA SERVIDOR.

```
spobras<- spobras %>% mutate(ANOS_DE_TRABALHO=(today()-ADMISSÃO)/dyears(1))
spobras$ANOS_DE_TRABALHO<-as.integer(spobras$ANOS_DE_TRABALHO)
spobras[,c(4,7,15)]
```

```
## # A tibble: 2,362 x 3
##   NOME                                ADMISSÃO   ANOS_DE_TRABALHO
##   <chr>                             <date>         <int>
## 1 WALTER MARTINS DE SOUZA          1976-08-19         44
## 2 LUIZ CARLOS DE SOUZA BARBOZA     1979-10-12         41
## 3 VALDELI DA SILVA                 1980-04-10         40
## 4 MARCIA APARECIDA TEIXEIRA DE CAMPOS CARVALHO 1980-07-14         40
## 5 MARCOS SILVA ROCHA                1981-08-03         39
## 6 JOAO PAULO RESPLANDES DE OLIVEIRA 1981-09-17         39
## 7 ELIANA MARIANO FRANCO             1981-12-01         38
## 8 SILVANA ISGROI CARVALHO           1981-12-28         38
## 9 MARCIA HELENA NUCCI EIRAS GARCIA  1983-11-01         37
## 10 ALFREDO LENZ                     1987-11-18         32
## # ... with 2,352 more rows
```

MEDIA SALARIAL e DESVIO PADRAO POR CARGO OCUPADO

```
spobras %>% group_by(CARGO) %>% summarize(media_salarial=mean(VENCIMENTOS,na.rm = TRUE),
                                          desvio_padrao=sd(VENCIMENTOS,na.rm = TRUE),Qtd=n()) %>% arrange(Qtd)
```

```
## 'summarise()' ungrouping output (override with 'groups' argument)
```

```
## # A tibble: 44 x 4
##   CARGO                                media_salarial desvio_padrao   Qtd
##   <chr>                             <dbl>         <dbl> <int>
## 1 DIRETOR DE OBRAS                    30062.         23125     4
## 2 CHEFE DE GABINETE                  24460.         10622    12
## 3 GESTOR SENIOR                       22534.          6541    24
## 4 DIRETOR PRESIDENTE                 22484.         12925    13
## 5 DIRETORA DE REPRESENTAÇÃO DOS EMPREGADOS 22354.         13351    12
```

```
## 6 DIRETOR ADMINISTRATIVO E FINANCEIRO      22327.      15548.      13
## 7 DIRETOR DE PROJETOS                      21293.      12262.       5
## 8 COORDENADOR DE NUCLEO                    19179.       6950.     162
## 9 COORDENADOR DE AREA PLENO                19032.       4347.     24
## 10 DIRETOR TECNICO                         18500         0       8
## # ... with 34 more rows
```

MAIOR SALARIO MENSAL RECEBIDO ENTRE TODOS OS SERVIDORES

```
spobras %>% select(NOME,ANOS_DE_TRABALHO,VENCIMENTOS) %>% arrange(desc(VENCIMENTOS))
```

```
## # A tibble: 2,362 x 3
##   NOME                                ANOS_DE_TRABALHO VENCIMENTOS
##   <chr>                                <int>         <dbl>
## 1 MAX YAMATOGUE FAGUNDES                3          79028.
## 2 PAULO SANTORO DE MATTOS ALMEIDA        3          70634.
## 3 SELMA BARROS DOS SANTOS DIAS           3          64750
## 4 GIOVANI OLIVEIRA DA COSTA              2          64750
## 5 MAURICIO BRUN BUCKER                   2          60802.
## 6 VIVIAN PATRICIA BANDEIRA NOGUEIRA PINTO 2          41531.
## 7 MARINA A KYRIOPOULOS                  3          40586.
## 8 HILDA MITIKO IUAMOTO PACHECO           3          39683.
## 9 HILDA MITIKO IUAMOTO PACHECO           3          39139.
## 10 NATALIE MATA                          4          38568.
## # ... with 2,352 more rows
```